APPENDIX A:

EPA ENTERPRISE ARCHITECTURE CORE COMPONENTS

<u>**Appendix A**</u>

EPA ENTERPRISE ARCHITECTURE CORE COMPONENTS

<u>**The Exchange Network**</u>

The Exchange Network is the Environmental Protection Agency's (EPA) proposed approach for the exchange of environmental data among EPA, states, and other parties with whom EPA and states exchange information. The Exchange Network "vision" is to promote access to and exchange of quality environmental data while reducing reporting burden and increasing the efficiency of data exchanges between Exchange Network partners - the parties that officially participate in the Exchange Network. During the early Exchange Network implementation phase, "Exchange Network Partners" will include EPA, states, tribes, and territories. In the future, the term "Exchange Network Partners" will likely include other governmental and possibly non-governmental parties. The Exchange Network will gradually replace the traditional approach to information exchange that requires states to feed data directly into multiple EPA national program systems. These arrangements often vary from state to state, region to region, and program to program. The Exchange Network will also facilitate transparent and secure data exchanges that support specific analyses, such as the use of indicators for measuring environmental results. While Exchange Network participation is voluntary, EPA and states expect participation in the Exchange Network to become the preferred method for routine intergovernmental transfers of environmental data.

The Exchange Network consists of both technical and organizational frameworks. The organizational framework consists of the decision-making and operational structures for building, maintaining, using, and evolving the Exchange Network. The technical framework encompasses the hardware, software, and protocols, and related technical decisions needed for Exchange Network implementation

The Exchange Network uses the Internet and Internet-based protocols to standardize and streamline the information exchange process, and consists of nodes that support the exchange of data among Exchange Network Partners. The data exchange on these nodes will be formatted according to agreed upon, standardized Data Exchange Templates (DETs) that rely on common, Internet-based protocols. The DETs depend on data standards that represent documented agreements on quality, consistency, formats, and definitions of commonly shared data. The suite of DETs will be compiled and tracked in the Exchange Network Registry/Repository.

The data exchanges among Exchange Network partners are governed by Trading Partner Agreements (TPAs). TPAs specify the appropriate DETs and explicitly define the quality, timeliness, and format of the data. Initially, "Exchange Network partners" will include EPA, states, tribes, and territories. In the future, the term "Exchange Network Partners" is likely to include other governmental and possibly non-governmental parties.

As of December 2002, two States, Nebraska and Mississippi, have entered into TPA's with EPA. TPAs and DETs are discussed further below.

Exchange Network Rationale

The Exchange Network is both a strategic and collaborative approach between EPA and its Exchange Network partners intended to address the following trends in environmental information management:

- **Growing Complexity and Volume of Data** - As the scale and complexity of environmental challenges (and their associated data) grow, environmental managers will collect, assess, and securely exchange more data.

- **Evolving State/EPA Roles** - The devolution of environmental management from the federal to the state and local levels, and the attempts to use more "integrative" or "adaptive" management approaches, has dramatically broadened the universe of data and data exchange.

- **Increased Need for Integration** - Integrated environmental management requires integrated environmental information and nearly always requires information integrated across media, program areas, and geographic, political, and organizational boundaries.

- **Growth of the Internet and e-Commerce** - The Internet and its associated technologies are transforming information management approaches. They are also increasing public expectations for data access and presenting information security issues of a new magnitude.

On an individual basis, EPA's partners are responding to these trends by making major investments in their internal, often integrated, systems. As part of these investments, many states have been supplementing their use of EPA national systems. EPA is currently in the process of developing its first Agency-wide Enterprise architecture and SDWIS is being used as pilot program data system to help establish options, costs and benefits to integrate with the Enterprise Architecture. While individual Partner efforts are important, there is also a need for a clear vision or framework for how Partners' systems will inter-operate and collaborate. Such collaboration and the data flows that support them are essential to meeting current and future environmental challenges.

In the future, managing these interchanges on a system-by-system, program-by-program basis will not be sufficient to meet the identified trends in information management and needs. Without a common framework, it is likely that individual Partners will build better and faster, but incompatible systems and a tremendous opportunity will have been lost.

The Exchange Network is intended to be such a framework to provide the vision of how these systems will work together.

EPA's Node - Central Data Exchange (CDX)

The EPA has established a single portal on the Web for environmental data entering the Agency. The Central Data Exchange (CDX) offers companies, states, tribes and other entities a faster, easier, more secure reporting option. CDX provides built-in data quality checks, web forms, standard file formats, and a common, user-friendly approach to reporting data across vastly different environmental programs. A cornerstone of EPA's e-government initiative, CDX currently accepts data for certain air, water, waste and toxics programs and will gradually expand to support all Agency environmental reporting by 2004. Although its current focus is electronic, CDX will eventually incorporate a facility that centralizes paper data collections as well. CDX is part of a broader effort by states and EPA working together to build an Exchange Network to

integrate state and federal environmental data, reduce the burden of reporting, and improve data quality.

CDX benefits to reporting entities include:
- Reducing their reporting burden and associated costs.
- Enabling automated, machine to machine transactions eliminating tedious paper forms and redundant data entry.
- Ensuring a secure electronic environment.
- Improving data quality through built-in edit and data quality checks.
- Offering faster, easier click-and-send reporting with one consistent point of entry for reporting, one streamlined set of procedures, and one password.
- Confirming EPA's receipt of their data.
- Translating and distributing incoming data to the appropriate data system.

CDX benefits to EPA include:
- Centralizing receipt, security, user authentication, archiving, translation, distribution and related user support services for incoming data.
- Eliminating redundant infrastructure and its associated cost.
- Enabling the Agency to streamline and simplify compliance reporting for everyone.
- Establishing EPA's presence on the Exchange Network.
- Laying the groundwork for future data integration and quality improvement efforts with the States.

Existing Partner Information Systems

Exchange Network Nodes use two kinds of software to interact with back-end systems. Middleware maps the location, type, and format of data in the back-end systems with the type and format required for the XML schema. Database connectivity tools communicate between the Middleware and the database that houses the partner's data.

Partners will map their existing data to the agreed-upon data exchange templates (DETs) using their Middleware product. Mapping consists of identifying the location of the data in the back-end database, defining the format of the stored data, and defining the format of the output data (XML schema). Once the source data and the output data have been defined, the Middleware translates from source to output and back.

Partners can participate in the network regardless of their existing system architecture using these standard tools. Stand-alone databases, data warehouses, integrated databases, and enterprise integrated systems all can be connected to nodes. While it will be easier to connect a smaller number of systems to the node, any stable system that serves as a source of quality data can be used.

Exchange Network partners that have integrated their systems already will be especially well positioned for these connections. However, given the incremental nature of both integration and flow development, it is likely that most partners will connect a number of (non-integrated) systems to their nodes.

Existing technical architecture will determine the specific approach Exchange Network partners will take when connecting their nodes to their existing systems. Processes for update

schedules for databases and warehouses, back-up schedules, and quality control timing will all influence how and when nodes can access data. While logically straightforward, mapping the Middleware to the existing systems is not trivial; it will require planning and staff time.

Data Standards

Data standards are the documented agreements on data formats and definitions of common data. Data standards are especially important tools for data integration and exchange because they allow data from many compliant sources to be integrated. The benefits of data standards are even greater for Exchange Network partners because they reduce ambiguity of the information contained in DETs at the most rigorous level possible. Standards are especially important for large-scale integration and aggregation efforts such as those performed by EPA.

Data Exchange Templates

 DETs describe and enforce the format and specific restrictions, where applicable, of the data being exchanged across the Exchange Network. Specifically, DETs are either XML Document Type Definitions (DTDs) or XML schema. Exchange Network implementation requires not only that these DETs be developed and used, but also that their development and coordination be harmonized to ensure compatibility across network flows. DETs will continue to be developed as new data standards arise and existing standards are improved. Used together, the data standards and DETs will provide partners with powerful tools for data access and integration.

Trading Partner Agreements

Trading Partner Agreements (TPAs) are documents that Exchange Network partners agree upon for each flow. They define what flow(s) are exchanged, define the stewardship and security expectations, and specify additional technical details for the exchange of information among two or more Exchange Network partners. A TPA is, or can be defined as, a stand-alone document, an addendum or supplement to an existing agreement, or part of an existing agreement. If existing agreements and their amendments satisfy the minimum set of elements that document the content and process of a data flow, then a separate, stand-alone document is not required. For the purposes of this Plan, all such agreements are called TPAs.

Exchange Network partners will need to develop at least a basic internal strategy for managing multiple TPAs across programs and with various offices and agencies. The strategy should address priorities for Exchange Network flows to be documented in TPAs, resource and staffing issues and implications for current business and management processes associated with data exchange.

Stewardship

The flow of quality data is fundamental to the Exchange Network. The concept of stewardship refers to the responsibility for this data quality on the Exchange Network. Data partners will take responsibility for the data they place on the Exchange Network and for their interactions with the Exchange Network itself. These responsibilities will be spelled out in Trading Partner Agreements. The concept of stewardship is involved in each of the components of the Exchange Network. Two of the most important of these are Data Stewardship and Node Stewardship.

Data Stewardship - By agreeing to host and exchange data and information, each trading partner on the Exchange Network assumes and accepts certain data stewardship responsibilities:
- Assuring that responsibilities for data quality and integrity are clearly defined and understood inside the organization.
- Assuring that data source, derivation and accuracy meet specifications.
- Assuring that data formats and units of measure meet specifications.
- Assuring that any other relevant data or metadata meet the specifications in the TPA.

Node Stewardship - Each partner, whether state, tribal or federal will be the steward of its own node, making sure that it functions properly and that the data available complies with agreed upon terms:
- Assuring that the hardware and software that create, manage, store and provide access to the data work properly.
- Assuring that the data transmitted and received is complete.
- Assuring that the data transmitted and received comply with agreed-upon formats and time schedules.
- Assuring that data has not been altered.
- Assuring that confidential and sensitive data has not been intercepted.

System of Registries

The System of Registries (SoR) is a centralized data registry that provides an authoritative source of information critical to data integration and exchange between EPA and its partners. The SoR supports the Agency's environmental information network by uniquely identifying objects of interest to EPA including information resources, facilities, chemicals, biological organisms, and data elements. The SoR provides the means for coordinating the management, access, and use of EPA's core registry systems. Information is accessed through several registry systems including the Information Resource Registry System (IRRS), the Environmental Data Registry (EDR), the Terminology Reference System (TRS), the Substance Registry System (SRS), the Chemical Registry System (CRS) and the Biology Registry System (BioRS). The SoR also links to the Facility Registry System (FRS) and the Environmental Information Management System (EIMS)-two additional EPA core registry systems. Some of the key registries include:

The Facility Registry System (FRS) - The FRS is a centrally managed database developed by EPA's Office of Environmental Information (OEI). It provides Internet access to a single source of comprehensive information about facilities subject to environmental regulations or of environmental interest. The FRS contains accurate and authoritative facility identification records, which are subjected to rigorous verification and data management quality assurance procedures. FRS records are continuously reviewed and enhanced by a Regional Data Steward network and active state partners. The facility records are based on information from EPA's national program systems and state master facility records and enhanced by other Web information sources. The Central Data Exchange (CDX) registration, when fully implemented, will also be used to create and update facility identification records. As of July 2002, FRS has over 1,133,484 unique facility records linking to over 1,497,987 program interests.

The Facility Registry System also includes locational information that provides accurate mapping of the facilities regulated by EPA.

In terms of benefits, the FRS will:

- Reduce the long-term reporting burden for facilities, states and programs.
- Improve data quality by helping to reduce errors in state and Agency facility information.
- Provide better tools for cross-media environmental analysis.
- Provide better public access to the Agency's environmental information.
- Give facilities the flexibility to review and update their identification information.

The Office of Ground Water and Drinking Water has listed all of its community water systems in the FRS.

The Substances Registry (SRS) - Chemicals, Biological Organisms, and Miscellaneous Substances - SRS serves as the nucleus for linking information about substances regulated by the EPA. The SRS search page includes queries for substances (such as chemicals, organisms, and physical characteristics) in EPA regulations, data systems, and other information resources.

The Chemical Registry (CRS) - Chemicals with Corresponding Information Resources - CRS provides information on chemical substances and how they are represented in the EPA regulations and information systems. The CRS search page includes queries for chemicals by common identifiers.

The Biological Registry (BioRS) - Biological Organisms with Corresponding Information Resources - BioRS provides information on biological entities and how they are represented in the EPA regulations and information systems. The BioRS Search page includes queries for biological organisms.

The Environmental Data Registry (EDR) - The EDR is a comprehensive, authoritative reference for information about the definition, source, and use of environmental data. The EDR supports the creation and implementation of data standards that are designed to promote the efficient sharing of environmental information among EPA, states, tribes, and other information trading partners. The EDR also catalogs data elements in application systems. The EDR does not contain environmental data - it provides descriptive information to make the data more meaningful.

Exchange Network Registry/Repository - The Exchange Network Registry/Repository is a website that serves as the official record and location for the Exchange Network's DETs. The Registry/Repository will also store other Exchange Network documents such as TPAs. Trading partners will depend upon the Registry/Repository to access the templates to validate flows they receive and properly structure flows they are sending. The Registry/Repository will be used both manually by users to get copies of DETs for implementation, and automatically as nodes request DET information "on-the-fly" during the process of a data exchange. In addition, the Registry/Repository will be used to indicate the status of DETs, including their compliance with applicable standards, their acceptance by EPA and other information. The Registry/Repository will also provide an ideal way for parties interested in similar DETs to become aware of each other.

APPENDIX B:

OGWDW INFORMATION ARCHITECTURE CORE COMPONENTS

Appendix B: OGWDW Information Architecture Core Components

BASELINE ARCHITECTURE

The Federal CIO Council defines Baseline Architecture as – *"the set of products that portray the existing enterprise, the current business practices, and technical infrastructure, commonly referred to as the "As-Is" architecture."* This section describes in some detail the currently existing information architecture.

<u>Business Processes</u>

The following narrative provides an overview of the major programs implemented by the Office of Ground Water and Drinking Water (OGWDW) under the Safe Drinking Water Act (SDWA), their goals, purposes and objectives and a high level description of the major business processes engendered by these programs. Future iterations of this Information Strategy Plan will include more detailed business process flow diagrams of the major business processes as appendices.

The Safe Drinking Water Act and Program Management

Congress enacted the Safe Drinking Water Act in 1974 and has enacted major amendments in 1986 and 1996. The purpose of SDWA is to establish national enforceable standards for drinking water quality and to guarantee that water suppliers monitor water to ensure that it meets national standards. The 1974 SDWA restructured drinking water programs in two significant ways.

First, it set up a higher level of responsibility for regulating drinking water systems than established state programs by forming a federal program, called the Public Water System Supervision Program (PWSS).

Second, it expanded the focus from water system planning and prevention of contamination, to include developing standards, monitoring for contaminants, and taking enforcement action.

Federal law required the development of federal regulations. However the law realized that protection of drinking water was still primarily a state responsibility. SDWA included a major focus on delegating primary responsibility for program implementation (i.e., primacy) to the states.

EPA's Director of OGWDW is the National Program Manager of the SDWA. Accordingly, OGWDW develops national policy and sets national goals and priorities for drinking water programs. OGWDW consists of two divisions: the Standards and Risk Management Division and the Drinking Water Protection Division.

The Standards and Risk Management Division (SRMD) is responsible for setting drinking water standards and monitoring requirements, establishing priorities for new standards, and researching technologies that water systems can use to comply with new and existing standards.

SRMD includes the Technical Support Center. The Technical Support Center provides technical and scientific support for the development of drinking water standards and their implementation. In addition, it manages the implementation of the Information Collection Rule and the drinking water laboratory certification program, and supports the Partnership for Safe Water, treatment plant optimization and analytical methods development.

The Drinking Water Protection Division oversees implementation of SDWA regulations through various programs. They are: the public water system supervision, underground injection control (UIC), source water assessment and protection, sole source aquifer and wellhead protection programs. **It is also responsible for maintaining drinking water information through computer databases and the Internet,** administering the State Drinking Water State Revolving Fund, and promoting consumer awareness of drinking water issues.

Other EPA Offices also have responsibilities for implementing SDWA:
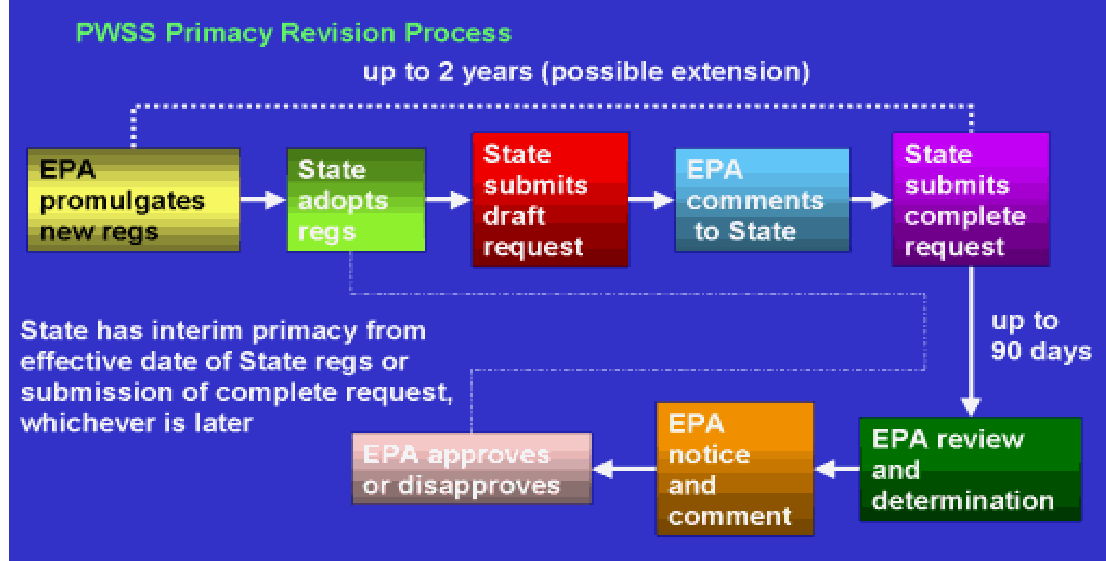
- The Office of Enforcement and Compliance Assistance enforces the statute and regulations;

- The Office of Research and Development is responsible for research related to health risk assessment, health effects, engineering and technology, monitoring, and quality assurance for drinking water issues; and

- The ten EPA Regional Offices implement drinking water programs in non-primacy states and provide liaison, coordination and oversight of the primacy states as defined below. In performing these activities, the regional offices perform inspections of water systems, provide implementation assistance to primacy agencies and water systems, take enforcement action where appropriate, administer the PWSS grants, and generally represent EPA interests with the state and local governments.

State Primacy

SDWA provides that EPA may delegate responsibility for implementation and enforcement of SDWA drinking water regulations to the states that meet the minimum federal requirements for the stringency of their regulations and the adequacy of their enforcement procedures. Primacy state programs operate in lieu of the federal drinking water program.

States and tribes must meet these requirements in order to obtain primary enforcement authority ("primacy") for the PWSS or UIC programs.

Concept of Primacy

PWSS Primacy Revision Process

up to 2 years (possible extension)

EPA promulgates new regs → State adopts regs → State submits draft request → EPA comments to State → State submits complete request

State has interim primacy from effective date of State regs or submission of complete request, whichever is later

up to 90 days

EPA approves or disapproves ← EPA notice and comment ← EPA review and determination

As EPA promulgates new regulations, primacy states must adopt the new requirements under state law and apply for primacy for those requirements. One important requirement is that the primacy agency provides inventory, violation and enforcement data to EPA on a regular basis. This data is stored centrally at the federal level in the Safe Drinking Water Information System (SDWIS-FED). Where a primacy agency fails to enforce regulatory requirements in a specified period of time, the SDWA requires EPA to initiate appropriate enforcement action. This is one of the major uses of data submitted by primacy agencies to EPA.

In states without primacy, EPA has primary enforcement authority. These states are called "Direct Implementation" or DI states because EPA directly implements the UIC and PWSS programs in those states.

Making changes to SDWIS-FED to accommodate regulatory changes and accommodating the primacy business process for adoption of new EPA regulations by the primary agency as shown in the above graphic is a major business process of the information management program.

The Public Water System Supervision Program

The public water system supervision program authorizes the regulation of the facilities that treat, store and distribute drinking water to taps; the PWSS program implements the National Primary Drinking Water Regulations developed and issued by EPA. The PWSS program also implements programs to enhance water system operation.

Public Water System Supervision Program

Over 167,000 Public Water Systems Nationwide

54,064

93,210

20,559

CWSs   NTNCWSs   TNCWSs
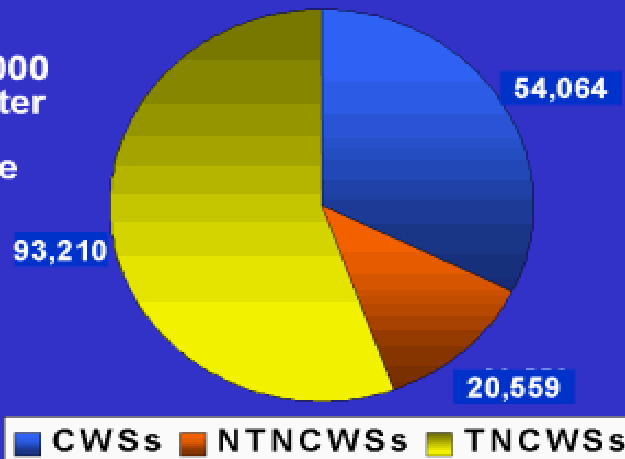
PWSs are divided into community water systems, transient non-community water systems (TNCWSs), and non-transient, non-community water systems because the risks to the populations these systems serve vary.

As shown above, the majority of PWSs are TNCWSs. While these systems are numerous, they do not serve the majority of the population because each system only serves a small number of people. However, almost everyone is served by transient non-community water systems at some point. (TNCWSs include roadside stops, commercial campgrounds, hotels, restaurants, and other facilities that have their own water supplies and serve a transient population at least 60 days per year). Community water systems serve the vast majority of the population. A community water system can be vast, serving millions of people (like New York City or Boston) or small, serving a trailer park with 25 residents.

There are currently over 160,000 water systems regulated by the Federal government in the U.S. National Primary Drinking Water Standards established either the maximum concentration of pollutants allowed in or the minimum treatment required for water that is delivered to customers. A Maximum Contaminant Level Goal (MCLG) is the maximum level of a contaminant in drinking water at which no known or anticipated adverse health effects would occur. A Maximum Contaminant Level (MCL) is enforceable. It is the maximum permissible level of a contaminant in water that can be delivered to any user of a public water system. An MCL is set as close to an MCLG as possible, taking into account the costs and benefits and feasible technologies.

For some contaminants, there is not a reliable method that is economically and technologically feasible to measure the contaminant, particularly at low concentrations. In these cases, EPA establishes a treatment technique. A treatment technique is an enforceable procedure or level of technological performance that public water systems follow to ensure control of a contaminant.

An example of a treatment technique involves protection of consumers from certain pathogens. Reliably measuring the concentration of pathogens can be cost prohibitive. EPA has found that operation of filters at a certain level of performance would reliably remove the pathogens from the water. EPA implemented regulations requiring filtration at a specified level of performance.

In the regulatory scheme provided by the SDWA, EPA conducts and/or analyzes public health research and other data regarding the public health impacts of a contaminant, evaluates treatment and control technologies and associated costs, conducts risk assessments on public health impacts of various levels of a contaminant, and establishes a MCL or treatment technique it determines is economically achievable. The EPA also establishes monitoring requirements for these contaminants which specify the number and types of samples to be collected, the frequency of sampling, sampling locations in the water system, the analytical methods to be used and related technical requirements. Public Water Supply systems have the responsibility for providing the necessary treatment or controls, conducting the necessary monitoring, and submitting monitoring results to the primacy agency. The primacy agency (usually states), have primary responsibility for determining compliance and taking necessary enforcement actions.

EPA Regional Offices oversee and track primacy agency state enforcement efforts and directly enforce the regulations in DI states. Oversight and enforcement focus on actions against significant non-compliers (SNCs). Significant noncompliance presents a potentially serious public health concern (as opposed to a single monitoring violation, for example).

The primacy agency submits certain information and data on PWS's and violations of regulatory requirements to EPA on a routine basis. EPA compiles this data, does quality control checks, analyses the data, calculates SNC's, where necessary provides compliance assistance to the primacy agency or public water supply or in some instances takes federal enforcement action to compel compliance. EPA also makes data available to the public, develops national trends and statistics, prepares formal reports to Congress and uses the data to assist in further policy or regulatory development. SDWIS-FED is the EPA information management system that supports this high priority business process.

EPA has developed SDWIS-STATE, an information management system designed to assist smaller states with limited or no automated information management systems of their own. SDWIS-STATE is of much broader scope and much larger than SDWIS-FED because it is designed to help the states manage the entire PWSS program including additional state program requirements. Information contained in SDWIS-FED is a small subset of information contained in SDWIS-STATE. Currently, 25 states, 6 EPA Regions and 2 territories are using SDWIS-STATE.

The Drinking Water State Revolving Fund Program

The Nation's 54,000 community water systems make significant investments to install, upgrade, or replace infrastructure to continue to ensure the provision of safe water to their 254 million customers. Installation of new treatment facilities can improve the quality of drinking water to comply with national primary drinking water standards and protect public health. Improvements are also needed to help those water systems experiencing a threat of contamination due to inadequate distribution and transmission pipes.

Many public water systems find it difficult to obtain affordable financing for infrastructure improvements that would enable systems to comply with national primary drinking water

standards and protect public health. Recognizing this fact, Congress established the Drinking Water State Revolving Fund (DWSRF) as part of the 1996 SDWA Amendments. The goal of the program is to provide states with a financing mechanism for ensuring safe drinking water to the public. States can use federal capitalization grant money awarded to them to set up an infrastructure funding account from which assistance is made available to public water systems.

Loans made under the program can have interest rates between 0 percent and market rate and repayment terms of up to 20 years. Loan repayments to the state will provide a continuing source of infrastructure financing into the next century. The program also places an emphasis on small and disadvantaged communities and on programs that emphasize prevention as a tool for ensuring safe drinking water.

Congress provided $1.275 billion for the DWSRF program in fiscal year 1997. The amount of funding each state was eligible to receive in 1997 was based on a formula used to award state program grants under the Public Water System Supervision program. Congress has provided an additional $3.145 billion for the DWSRF program for fiscal years 1998 through 2001, including $825 million for fiscal year 2001. The amount of funding each state is eligible to receive for fiscal years 1998 through 2001 is based on the total eligible need determined for each state by the Drinking Water Infrastructure Needs Survey which the EPA released in January 1997.

Both publicly and privately owned community water systems and non-profit non-community water systems are eligible for funding under the DWSRF program. Eligible projects include installation and replacement of failing treatment facilities, eligible storage facilities and transmission and distribution systems. Projects to consolidate water supplies may also be eligible.

States develop a priority system for funding projects based on three criteria from the Act. States rank the projects and then offer loans to systems based on their ranking order. Priority is given to those eligible projects that:

- address the most serious risk to human health;
- are necessary to ensure compliance with the requirements of the SDWA
- assist public water systems most in need according to state-determined affordability criteria.

The Drinking Water SRF National Information Management System collects information that provides a record of progress and accountability for the program. The system is managed by OGWDW and data is made available to the public on the World Wide Web.

The Underground Injection Control (UIC) Program

Underground injection is the technology of placing fluids underground, in porous formations of rocks, through wells or other similar conveyance systems. This technology is used for many purposes including disposal of wastes and oil recovery. While rocks such as sandstone, shale, and limestone appear to be solid, they can contain significant voids or pores that allow water and other fluids to fill and move through them. Man-made or produced fluids (liquids, gases or slurries) can move into the pores of rocks by the use of pumps or by gravity. The fluids may be water, wastewater or water mixed with chemicals. Injection well technology can predict the capacity of rocks to contain fluids and the technical details to do so safely. Facilities across the United States discharge a variety of hazardous and non-hazardous fluids into more than 400,000 injection wells.

The Safe Drinking Water Act established the UIC Program to provide safeguards so that injection wells do not endanger current and future underground sources of drinking water (USDW). The most accessible fresh water is stored in shallow geological formations called aquifers and is the most vulnerable to contamination. These aquifers feed lakes; provide recharge to streams and rivers, particularly during dry periods; and serve as resources for 92 percent of public water systems in the United States.

The UIC Program defines an injection well for a wide variety of injection practices that range from more than 100,000 technically sophisticated and highly monitored wells which pump fluids into isolated formations up to two miles below the Earth's surface, to the far more numerous on-site drainage systems, such as septic systems, cesspools, and storm water wells that discharge fluids a few feet underground.

The EPA groups underground injection into five classes for regulatory control purposes. Each class includes wells with similar functions, and construction and operating features so that technical requirements can be applied consistently to the class.

Benefits of the UIC Program

Injection wells have the potential to inject contaminants that may cause underground sources of drinking water to become contaminated. When wells are properly sited, constructed, and operated, underground injection is an effective and environmentally safe method to dispose of wastes. The goals of the EPA's UIC Program are: to prevent contamination by keeping injected fluids within the well and the intended injection zone, or, in the case of injection of fluids, directly or indirectly into a USDW; and, to require that injected fluids not cause a public water system to violate drinking water standards or otherwise adversely affect public health. These minimum requirements affect the siting of an injection well, and the construction, operation, maintenance, monitoring, testing, and finally, the closure of the well. Injection wells require authorization under general rules or specific permits. Finally, states may apply to have primary enforcement responsibility (primacy) for the UIC Program. To date, 33 states, Guam, the Commonwealth of the Mariana Islands and Puerto Ricohave obtained primacy for all classes of injection wells. Seven states share primacy with the EPA. The EPA administers UIC programs for the remaining states, the Virgin Islands, American Samoa and Indian Country.

At the present time, information management systems for the UIC program are scattered among the states, EPA regions and headquarters. Presently, a national schema or unified set of data management requirements does not exist.

The Source Water Protection Program

Source water is untreated water from streams, rivers, lakes, or underground aquifers that is used to supply private wells and public drinking water. Most public and some private well drinking water is treated before it enters homes. While some treatment is usually necessary, ensuring that source water is protected from contamination can reduce the cost of treatment and the risk to public health.

Most source water is defined as surface or ground water. The majority of drinking water in large metropolitan areas originates from a surface source such as a lake, stream, river or reservoir. The

land area that can have an impact on these water bodies is called a watershed, and can be delineated on a map.

Most water in smaller communities originates from under ground and is pumped to the surface through a well. Ground water comes from natural underground layers, often of sand or gravel, which contain water. These formations are called aquifers. The land area that can have an impact on the quality of this underground water is called the aquifer recharge area.

There are many contaminants that may be present in source water before it is treated.
These include:

- Microbial contaminants, such as viruses and bacteria,
- Inorganic contaminants, such as salts and metals,
- Pesticides and herbicides,
- Organic chemical contaminants, including synthetic and volatile organic chemicals,
- Radioactive contaminants.

Assessing the Risks

While many states, water systems, and localities have watershed and wellhead programs, the 1996 Safe Drinking Water Act Amendments placed a new focus on source water quality. States have been given access to funding and required to develop Source Water Assessment Programs (SWAP) to assess the areas serving as public sources of drinking water in order to identify potential threats and initiate protection efforts.
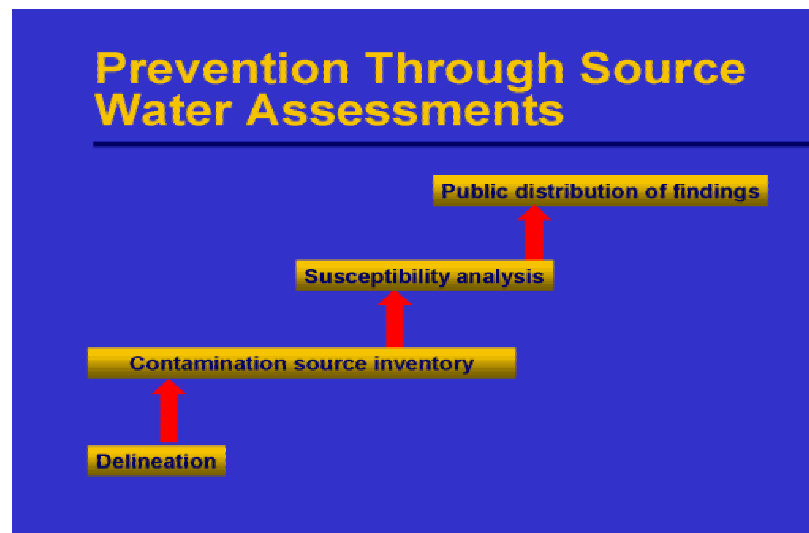
The source water assessment programs created by states differ since they are tailored to each state's water resources and drinking water priorities. Each assessment includes the four major elements shown above:

- delineating (or mapping) the source water assessment area
- conducting an inventory of potential sources of contamination in the delineated area
- determining the susceptibility of the water supply to those contamination sources
- releasing the results of the determinations to the public.

Benefits of the Source Water Protection Program

Protection of drinking water at the source can be successful in providing public health protection and reducing the treatment costs/challenge for public water suppliers. Source water quality can be threatened by many everyday activities and land uses, ranging from industrial wastes to the chemicals applied to suburban lawns. Private well owners are urged to test regularly for common contaminants such as microbes and nitrate-nitrogen because there is no federal oversight of their water source. Water systems are heavily regulated through the Public Water System Supervision Program and respond to this threat to public health with regular water quality monitoring and actions ranging from well closure to expensive treatment. In some cases, source water protection can eliminate or forestall the need to change or modify treatment processes. Treatment is expensive and source water protection can save consumers significant money.

Whether a public water system relies on surface water, ground water, or a combination of the two, protection of a water system's source is important. Prevention of contamination is one of the most cost-effective methods of ensuring safe drinking water supplies. If source water becomes contaminated, expensive treatment or replacement of the water source may be required before safe drinking water can be delivered to users. Treatment costs are passed on to every user served by the public water system.

Once completed, source water assessment results can be used to focus prevention resources on drinking water protection. EPA strongly encourages linking the source water assessments to implementation of source water protection programs. The Source Water Protection (SWP) Program is a non-regulatory program at the federal level.

At the present time, information management systems for the SWP program are scattered among the states, EPA regions and headquarters. Surface water sources of contaminants are contained in the Permit Compliance System (PCS) managed by the Office of Enforcement and Compliance. PCS is one of the largest public data systems in the nation. Presently, a national schema or unified set of data management requirements for all source waters does not exist.
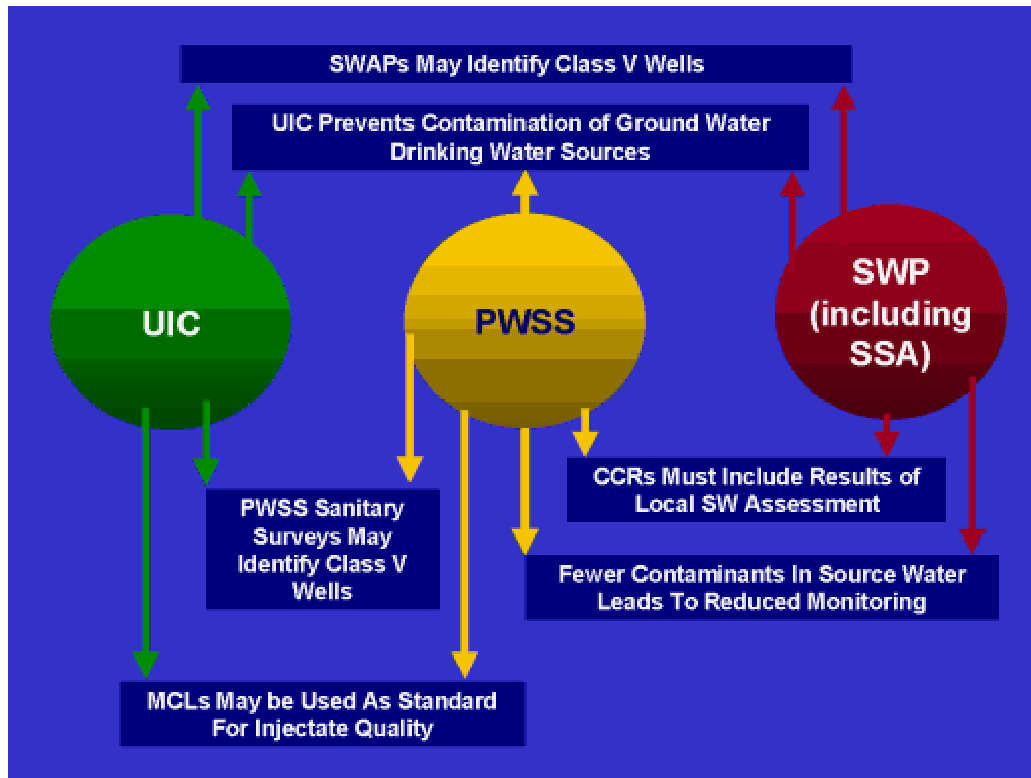
The three major programs within the Office of Ground Water and Drinking Water are interrelated in many ways. The common goal of PWSS, UIC, and SWP at the federal, state, and local levels is to protect public health. The graphic above shows just a few ways that the programs relate to each other. Integration of source water data at this time is very limited.

The Unregulated Contaminant Monitoring Program

The 1986 SDWA Amendments required EPA to establish a list of substances that were not regulated at that time but had the potential for adverse public health impacts and to conduct a national monitoring program at PWSs to determine their presence and concentrations in drinking water supplies. The Amendments required periodic revision of the list and re-sampling to be conducted at five-year intervals. Two rounds of monitoring occurred under this provision.

The Round 1 dataset contains public water system monitoring sample results for 62 (then) unregulated contaminants, generally collected between 1988 and 1992, from 40 states and primacy entities. Round 1 data were stored in a database called the Unregulated Contaminant Monitoring Information System (URCIS).

The Round 2 dataset (the second round of unregulated contaminant monitoring) contains public water system monitoring sample data for 48 (then) unregulated contaminants, generally collected between 1993 and 1997, from 35 states and primacy entities. Round 2 data were incorporated in the EPA Safe Drinking Water Information System, SDWIS/FED, that was modified to receive parametric data.

The monitoring for unregulated contaminants was conducted by the PWSs and sent to the state primacy agencies that forwarded the data to EPA for evaluation.

The 1996 SDWA Amendments modified but continued the unregulated contaminant monitoring program established by the 1986 Amendment. Under these Amendments, EPA issued the 1999 Unregulated Contaminant Monitoring Rule (UCMR 1999) which established a list of 12 contaminants to be monitored nationally to determine their presence in public water supplies.

Under EPA's Information Integration Initiative of FY 2000, OGWDW and EPA's Office of Environmental Information established a new database for UCMR. This database is able to receive data directly from large laboratories with sophisticated automated data entry, using an XML data format by way of the World Wide Web. This effort also developed Web forms for use by smaller PWSs for their entry of data and transmission over the Internet to EPA. EPA holds the data for a period of 60 days during which period states and PWSs are able to access the data over the Internet and submit comments to EPA on the results of their review. After this period, EPA is free to use this data for rule making and to provide public access to the data.

The National Contaminant Occurrence Database (NCOD)

National Contaminant Occurrence Database (NCOD) was developed in response to the 1996 SDWA Amendments. The data collected and stored in this database, like the unregulated monitoring data, is used to support EPA's decisions related to identifying contaminants for regulation and subsequent regulation development. The NCOD contains contaminant occurrence data for both regulated and unregulated contaminants in public water systems from PWSs and other sources such as the U.S. Geological Survey National Water Information System on physical, chemical, microbial and radiological contaminants. Regulated occurrence data are sample data from monitoring in public water systems for contaminants with health-based standards established by EPA under the SDWA.

EPA uses NCOD data and the data generated under the UCMR (1999) to evaluate and prioritize contaminants on the EPA Contaminant Candidate List (CCL). The CCL is a list of contaminants EPA is considering for possible new or revised drinking water standards.

<u>Data Model(s)</u>

SDWIS-FED
SDWIS-FED is designed to support OGWDW in monitoring compliance with the Safe Drinking Water Act. SDWIS-FED processes the following major categories of information:

- Characteristics of Public Water Supply Systems (includes administrative contact information, activity status, PWS type, population served, primary source type, and owner type).
- Water system facility and treatment data (includes flow data between sources of water through the treatment plant).
- Locational and geographic data to support geospatial applications and source water assessments.
- Violations of the national Primary Drinking Water Standards and other implementing regulations of the Safe Drinking Water Act.
- Enforcement and compliance assistance actions (formal and informal) and linkage data to violations.
- Sample data for unregulated contaminants.

The SDWIS/FED database is a 3rd normal form relational database, comprised of over 50 tables. Many of the tables are look-up tables as well as association tables. About 11 of the tables contain actual data (other than the look-up and association). There are several tables, and attributes within existing tables that are not populated, due to an initial intent of having a single data model for SDWIS/FED and SDWIS/STATE applications. This was determined to not be a feasible alternative, and the resources to remove the unused structure from the FED database were not expended.

A data warehousing model and associated on-line analytical processing (OLAP) capabilities now exist. Data is extracted and reformatted into the warehouse from the SDWIS/FED system quarterly update. A test data warehousing environment will be operational shortly. It will likely provide Intranet access in the short term. Production Internet access is still to be defined.

Inventory data is reported on an annual basis, while other data is reported quarterly with a one quarter reporting lag. The data is frozen each quarter after all processing and validation is completed.

SDWIS-STATE

SDWIS-STATE is a relatively new system developed to assist states that did not have automated information management systems or the capability of developing one of their own. SDWIS-STATE, as contrasted with SDWIS-FED, was designed to assist states in managing their entire drinking water program on a day-to-day operational level. SDWIS-FED's focus is on consolidating selected data as previously described on a national basis and making that limited set of data available to the states, regions and the general public.

SDWIS-STATE operates in a client-server platform using a UNIX operating system or one of several versions of Windows operating systems. It uses an Oracle database for the backend and the front end is written in C++. It contains 147 tables and 1,886 data elements. It addresses 726 analytes, 30 monitoring rules and 62 violation types. Currently, 25 states are using SDWIS-STATE. Data that is required to be reported to EPA is periodically extracted from SDWIS-STATE tables and converted to DTF format and submitted to SDWIS-FED where the data receives quality control checks and then entered into SDWIS-FED tables.

OGWDW Data Warehouse

Periodically, the data warehouse extracts SDWIS-FED data into staging tables modeled after SDWIS-FED tables. Additional QA is performed on the data and transforms it, adding attributes and de-normalizing the data, and organizing it by subject. Several data marts are also periodically updated which contain subsets of the data (in the form of multi-dimensional star-schema cubes), which facilitate making analysis tools in the form of OLAP cubes and pivot tables, as well as an array of standard reports.

The OGWDW data warehouse includes SDWIS-FED inventory (water system, water system facility, treatments, contacts, and locational data) and compliance data; samples datasets listed above; and results of PWS audits performed by states.

Fact tables include current violations, violations organized to facilitate trend analysis, analytical results from static and active sample datasets and data verifications findings. Conformed dimensions (which are basically the same as EPA Registries) facilitate information integration as they can be used with different fact tables.

NCOD/UCMR

NCOD consists of static datasets for Rounds 1 and 2 (then) unregulated contaminants, for the 6-year review of 16 regulated contaminants from a 15-state sample, and for the current flow of unregulated samples data via the Safe Drinking Water Accession and Review System (SDWARS) (UCMR 1999). The static datasets have gone through extensive quality assurance and been evaluated for national representativeness, documented in EPA analyses and they are available for download from the web.

SDWARS, the UCMR (1999) transaction database, is housed at RTP; data is sent to it directly from laboratories in a number of formats, including XML. UCMR data from SDWARS is extracted into a data warehouse. Pivot tables are created to facilitate access.

APPLICATIONS

SDWIS-FED
<u>Reporting Toolkits</u>
The reporting of public drinking water inventory and noncompliance information to SDWIS-FED is supported by a variety of individually-developed state data systems as well as a Personal Computer (PC)-based, EPA-developed data entry tool (DTFWriter). A full-featured local database application (SDWIS-STATE) was developed by EPA for use by Primacy Agencies is also available. Due to complexity and pending obsolescence of DTFWriter, EPA decided to develop Actions DTF, a short-term, stand-alone, single-purpose, PC-based application that supports violations and enforcements data entry.

DTFWriter was developed using Clipper™. The system can run on any computer supporting PC/MS DOS version 3.0 or higher.

Actions DTF was developed to assist state and regional PC users in the creation of a data file containing Violation or Enforcement actions information that can be input to the SDWIS-FED System. The software creates records in DTF that is required for entry of data into SDWIS-FED. DTF files are input to the SDWIS-FED national database on a quarterly basis from the Primacy Agencies (states and EPA regions) that have been delegated PWSS oversight responsibility by EPA.

Actions DTF is a Microsoft™ (MS) Access® Windows application installed on a PC at the user site

SDWIS-FED contains a number of other applications.

<u>Data Entry subsystem</u> - This batch software (CLIST, JCL, CoolGen, COBOL, SAS, Assembler) performs input data editing and validation, constructs "total replace" transactions, posts data to the SDWIS-FED database, identifies, aggregates and creates error reports, and provides detailed and high level summaries of update status. Users are required to post the data to the EPA mainframe, and communicate data processing instructions to SDWIS-FED production control staff.

<u>Data Retrieval subsystem</u> - This is the software that creates the user interface for canned retrievals of data from the SDWIS-FED database. There are over 15 standard reports designed

for interactive batch access; storage of reports online or printed on high speed printers; or, provides access to the Platinum Report Facility, an *ad hoc* data retrieval tool.

SNC/Exception Tracking System - This software provides support to EPA's enforcement and enforcement oversight programs via generated SNC and exception records, three standard reports, and an on-line system for evaluation of noncompliance and enforcement data to allow regional modifications of the standard reports.

On-line Data Dictionary - This MS-ACCESS application provides the data dictionary for the database.

Error Code Database - This MS-ACCESS application provides a look-up for users debugging error reports to assist in understanding the nature of data entry errors and actions that need to be taken to correct those errors.

Data Warehouse - EPA staff operate, update and maintain a local data warehouse for data distribution and reorganization to enable easier access to SDWIS-FED data. Extract-transform-load (ETL) tools and procedures are utilized to extract data, transform it, and post it to the warehouse.

OGWDW Data Warehouse

There are two ways of accessing drinking water violations and inventory data:
- Through the mainframe—standard reports or *ad hoc* queries using PRF, as well as through use of the Oracle Transparent Gateway (OTG), and
- Through the OGWDW data warehouse

MS-ACCESS custom queries provide access to the warehouse tables and many of the data marts, as well as through several pivot tables, which can be downloaded off the web.

Numerous analysis tools in the form of pivot tables and OLAP cubes have also been built and are continually refined. These include:

- GPRA, violations, and inventory analysis tools for trends analysis.
- Current violations and inventory (including contacts, locational data, treatments, etc.).
- An array of data quality analysis tools based on both data verifications and SDWIS-FED data that assess data quality, completeness and accuracy of violations data, % of correct compliance determinations, rule implementation, timeliness of violations reporting, completeness of various required inventory elements including the Source Water Treatment Rule (SWTR) reporting and locational data.
- Several samples analysis tools for UCMR, 6-year review, and Rounds 1 and 2 datasets.

TECHNOLOGY

SDWIS-FED

The SDWIS-FED Reporting System is designed to operate on the IBM mainframe computer system; the data are held in an IBM Database2 (DB2) database.

The SDWIS-FED operating environment incorporates use of the following software:

- IBM's Interactive System Productivity Facility (SIPF).
- IBM's DATABASE 2™ (DB2) Relational Database Management System (RDBMS).
- Platinum Technology's Platinum Report Facility™ (PRF).
- User dialogue screens implemented using IBM's Dialog Management Services (DMS).
- Control processing via IBM's Time Sharing Option (TSO) Command Lists (CLISTs).
- Report production performed through a combination of original COBOL programs and COBOL programs modified to utilize SQL (Structured Query Language) formulated from user-supplied selection criteria.

EPA headquarters staff accesses the IBM mainframe via TCP/IP (Internet Protocol) -based communications between desktop devices and servers. EPA's 10 regional offices and state primacy sgency staff access the mainframe system through the Internet using *IBM WebSphere Host On-Demand™*.

SDWIS-STATE

SDWIS-STATE uses client-server architecture and supports Oracle, MS SQL Server and IBM's DB2 database system as well as several operating systems including UNIX, WindowsNT, Windows 98 and Novell. The servers are housed at the state primacy agency offices and EPA provides SDWIS-STATE software.

OGWDW Data Warehouse

The data warehouse is in a SQL Server database. The ETL tool is Microsoft Data Transformation Services (DTS). Several multi-dimensional OLAP cubes using MS Analysis Services software are available.

Data access for *ad hoc* queries is accomplished through MS-ACCESS databases, which have links to both SQL Server data warehouse tables, data marts, and SDWIS-FED mainframe tables. The Oracle Transparent Gateway is the means to access those tables that have not yet been pulled into the warehouse.

TARGET ARCHITECTURE

Business Processes

The EPA Target Business Reference Model as presented in the document entitled *"EPA Target Environmental and Health Protection Architecture"* (EHPA) developed by the Office of Environmental Information presents EPA's model for information integration. The Model for Information Integration (M4I)[1] was developed by the Information Integration Program and accepted by the Agency in July 2002. The M4I is a technical, strategic framework that proposes an integration of data, applications and technology across the Agency. It consists of the following high-level functions:

- **Connect and Exchange** — Electronically connecting to transmit or access data

---

[1] *Model for Information Integration, A Preview of the Core Components of the EPA's Target Environmental Information Architecture (EIA)*, July 24, 2002.

- **Process and Stage** — Data collection, cleansing, validation and approval for use
- **Store for Use** — Data storage, linkage and/or referencing for access and use
- **Use** — Data manipulation (potentially from multiple sources) to aid in learning, discovery and problem solving

Classifying major functions into these broad categories is intended to enable program and system managers across the Agency to think of information integration in general terms and to use common terminology to discuss and plan for their programs' functional needs. Classifying systems by common functions helps identify areas where improvements to services as well as reductions in costs can be made by eliminating redundancies through the sharing of services. OGWDW intends to employ this high level classification of functions as it further refines its planning in support of system modernization that not only meets its immediate programmatic business needs, but fully supports the enterprise business needs through conformance with the Agency's enunciated Enterprise Architecture.

The EHPA further lists and defines the following EPA business categories and subcategories:

- **Environmental Protection Services**
  - **Pollution Prevention**
    This area includes the Agency's non-command and control approaches to reduce or avoid pollution are centered, as well as its international voluntary efforts in such areas as ozone protection and climate change. Pollution prevention incorporates the current pollution prevention program, including such activities as the Design for Environment Program, the Energy Star program, waste minimization and a variety of best practice efforts. Also included are the pollution prevention aspects of regional programs such as Great Lakes and Chesapeake Bay.
  - **Pollution Control & Public Health Protection**
    This area includes the Agency's national standard-setting programs, such as those for ambient air quality and drinking water quality. It also includes source and facility permitting activities and other authorizations, along with supporting enforcement and compliance responsibilities. This business area relies heavily on state involvement. Under the Criminal Enforcement activities of OECA, this area supports homeland security through its environmental investigations and forensics functions.
  - **Emergency Response and Remediation**
    All cleanup operations, including Superfund sites, facility spills, transportation accidents, industrial accidents, oil spills, and other accidental releases of contaminants fall under this area. It is field-engineering oriented, under headquarters or regional office supervision. This area also supports homeland security responses, such as the anthrax decontamination of the U.S. Capitol and the World Trade Center response in New York.
  - **Environmental and Human Health Assessment**
    This area includes the Agency's responsibilities to monitor and evaluate current and future environmental conditions and human health risks. Encompassed here are activities that document, map and project many kinds of environmental trends. It is at this level that activities such as the proposed EPA Situation Room are found. These involve the integration of the Agency's knowledge of all dimensions of human health and environmental quality, and many are driven by the use of environmental indicators.

- **Shared Business Support Functions**
  - **Research and Development**
    EPA's Research and Development program supports the full range of top-level business areas. It conducts both basic scientific research as well as targeted research to support specific program needs. Areas include environmental studies (non-human biota and ecosystems), human health studies, development of monitoring and modeling methods, and creation of methods, standards and procedures to ensure the quality of scientific technical results. Under this general heading are the activities of the Agency's formal scientific and technical panels.
  - **Assessments**
    Grouped under this heading are analytical activities, such as risk assessment/risk management studies, economic impact analyses, social impact evaluations and legal reviews. Also within this area are the generation of environmental indicators: specification of methods to evaluate the state of the environment and to quantify relationships among Agency activities and their environmental results. These indicators directly support the business driver of supporting performance-based environmental protection.
  - **Regulatory Process Management**
    The Agency's regulatory processes include rulemaking activities, but also include the development of guidance documents and other activities in which public comment is invited or required. Activities within the process include the development of the rules themselves, the process of external review and comment, formal promulgation and the development of formal policy and guidance documents to facilitate implementation.
  - **Information Management**
    Management of information in its various forms includes: business-related information exchange from inside and outside the Agency; the processing of that information to conform to Agency systems; management of metadata standards governing Agency program data; data quality management operations to ensure the proper applications of standards to EPA data; integration of data to some form of enterprise repository; activities to ensure data security (data integrity, confidentiality and access); and, activities to deliver data in appropriate form to Agency personnel, public EPA partners, stakeholders and regulated parties. This area supports expectations for E-government and the need for better services over the Internet for stakeholders and the public.
  - **Communications and Training**

- **Program Management**
  This level of the target business architecture hierarchy covers activities that guide and direct activities at the program level. It includes program planning and design, formal delegations of authority under regulatory programs, partnership development, program implementation, and program analysis to determine effectiveness in relationship to goals and objectives.

EPA's target business architecture foresees the Agency's future as one in which quick access to authoritative and unambiguous information is essential. It is also one in which relationships among data—particularly the ability to draw clearer connections between program outputs and

environmental and public health protection—should be documented and made active in new or revised applications. EPA's target business model is characterized by highly interrelated functions that will ultimately rely upon highly integrated multimedia information to operate efficiently.

The model emphasizes the new focus on pollution prevention. This area will receive increased attention in the future as the Agency works to emphasize and implement the increased efficiency and cost effectiveness of preventing a problem versus fixing it after it happens.

Review of OGWDW's baseline business architecture above shows that OGWDW's business processes cover the full range of categories specified above. The following brief examples illustrate this fact:

Pollution Prevention:
OGWDW's Source Water Protection program is specifically designed through its Sole Source Aquifer and Wellhead Protection Programs to prevent contamination of groundwater sources of drinking water.

Pollution Control & Public Health Protection:
Development of Drinking Water Standards (MCLs) for protection of public health.

Emergency Response and Remediation:
Public health advisories are issued when public water supplies are known to be contaminated.

Development of Vulnerability Assessments of public water supply systems and appropriate responses to those vulnerabilities are receiving high priority in support of the newly emerging Homeland Security program.

Environmental and Human Health Assessment:
Both the ongoing National Contaminant Occurrence Database and Unregulated Contaminant Monitoring Program are aimed at assessing the potential for adverse human health impact of regulated and unregulated contaminants, and at providing data necessary to determine the need to develop or revise standards or otherwise regulate the release of these substances into the environment.

Similarly the shared business support functions and program management support functions can be mapped to the OGWDW baseline business architecture as presented previously in this document. The Agency's Enterprise Architecture Team is in the process of selecting tools that will enable the consistent mapping of program information into the Enterprise Architecture. A system called METIS may be the prescribed tool and will be adopted when it is fully supported by the Agency and available to the programs. It is OGWDW's intent to conform its target business architecture to the Agency's framework as outlined above.

While it is expected that many of OGWDW's business processes, as described in the high level presentation in the baseline business process discussion, will remain the same over the next 3 to 5 years, EPA will revise several business processes to meet the expanded information needs of OGWDW, which generally fall in the following categories:

- State oversight and assistance, including enforcement oversight

- National program oversight—key measures of program success, and program assessments

Information to the public

- EPA research, including developing and evaluating regulations
- Other needs including Homeland Protection, capacity development efforts
- Conformance to Agency's evolving Enterprise Architecture requirements including:
    - Participation in the Exchange Network through adoption of XML as the data transfer language between states and EPA
    - Use of the System of Registries particularly the Facilities Registry, Chemical Registry, Biological Registry, proposed XML and Metadata Registries
    - Use of the Central Data Exchange as OGWDW's data portal
    - Continued development of the OGWDW data warehouse and use of the Agency's data repository
    - Application of EPA data standards
    - Development of Trading Partner Agreements

There are many unmet needs in the existing information and business processes, and many opportunities to meet these needs more directly and effectively. EPA requires:

- Sample data on regulated contaminants, over time, in order to better evaluate existing regulations and develop new ones, to perform research on effects of multimedia pollutants, and to evaluate the success of the drinking water program;

- More effective and direct processes for conducting enforcement oversight;

- More meaningful and accurate measures of program success, to evaluate regions, various program initiatives including capacity development programs, infrastructure loans and drinking water resource security;

- More meaningful and accurate information to provide to the public.

- Optimization of data verification audits for:
    - State oversight—currently used to determine how well states are determining compliance with the regulations and reporting violations and inventory data to SDWIS-FED. Evaluate ways to include simple evaluations, and subjective assessments of state enforcement programs and capacity. Preserve statistically representative sample at the state-level.
    - National measures—need to optimize data verifications to also provide statistically-representative samples at the national-level, stratify samples across system size categories and ground water/surface water, in addition to water system type to most effectively evaluate:
        - SDWIS-FED data quality
        - Impact of data quality on the current GPRA measure
        - Investigate the possibility of replacing the current GPRA measure, which is based on reported violations, with statistically representative samples from audits.
        - Rule implementation

- Evaluate the implications of expanding the number and breadth of data verification audits.

- Evaluate ways to provide more timely and complete enforcement oversight
    - Calculate SNCs on PCs
    - Evaluate queries that can gather violations data from state data systems, calculate SNCs on-the-spot to enable QA and follow-up during visits to states.
    - Determine if there are ways to simplify and streamline the software used to calculate whether water systems are significant non-compliers

- Investigate the methods and program implications of obtaining data from sanitary surveys

- Investigate the methods and program implications of obtaining parametric data on regulated contaminants over time
    - Data flows—from states, from labs, through SDWARS?
    - Program benefits
    - Minimizing the misinterpretation and misuse of the data

- Investigate ways to improve the public's access to drinking water information via CCRs, enabling us to rely less on incomplete violations data

- Integrate drinking water data from several sources

- Integrate information
    - UIC and SRF
    - Integrate drinking water information across OGWDW and EPA

- Ways of integrating:
    - Geospatial tools
    - Data warehousing techniques
    - Agency Enterprise Architecture initiatives
        - Conformed dimensions/registries
        - Repositories

OGWDW Data Warehouse Processing and Access:

Pull all data fields into the warehouse from SDWIS-FED, which will be included in the future XML flows, warehouse them, and build an array of access tools from them. Information access shouldn't change when the new data flow through CDX begins.

Internet access: work with Envirofacts to modernize and replicate the mainframe standard reports, built from warehouse tables OGWDW provides them. Integrate the warehouse tables and data marts into the central repository and registries, etc. using ETL tools when they're ready for us.

Intranet access: post warehouse tables and some data marts that supply pivot tables and standard reports on an OW NT intranet server on the EPA Tree. This will be used as an access server.

And, of course, move the processing and warehouse storage from the mainframe to the NT server on the VLAN.

Next steps

Once new staging tables are built and modeled after the XML objects they will be populated with SDWIS-FED, replacing the current staging tables.

The SETS mainframe tools will be replicated and streamlined to run off the server or a PC

The possibility of loading DTF directly into the warehouse, bypassing SDWIS-FED, will be explored.

With these steps taken, the mainframe can be phased out of the SDWIS data flow.

Data Model(s)

SDWIS-FED

Transitioning from to an architectural environment employing use of staging tables and other data warehousing technology will have a major impact on the current data model/structure of SDWIS-FED including the ultimate elimination of the system as it now exists. In the near term, the transference of data edits/verification to states and EPA Regional Offices could entail some structural changes as a consequence. Also, the replacement of DTF with XML will potentially involve structural changes to SDWIS-FED. There will likely be a period of operational overlap between SDWIS-FED and these new data structures until the new systems are fully functional in the operating environment.

SDWIS-STATE

Transitioning from DTF to XML as the data exchange language between states and EPA and the data Edit/Validation responsibilities as described above will likely result in some structural modifications to the current SDWIS-STATE system to accommodate these changes. Full Web enabling of SDWIS-STATE (beyond use of XML as the data exchange language) will entail other structural adaptations.

UIC, SWP, SRF and GPRA

Information management systems for these programs will be reviewed to determine the potential for (and benefit of) development of national standards, expansion of these systems to satisfy unmet or evolving needs (including Homeland Security, Web and geographically enabling these systems), and conformance with Office of Water and EPA Enterprise Architecture policies and requirements (including consolidation and integration, where practicable).

Tracking GPRA requirements is clearly an Enterprise level activity and OGWDW will adopt any software that the Office of Water or the Agency develops and provides. In the interim, the OGWDW data warehouse will continue to support this need. The warehouse has been specifically designed to accommodate any course of action the Office of Water or the Agency takes in this regard.

Applications

SDWIS_FED
By 1/2004, a new application will be ready for distribution that is designed to run on local desktops and/or servers and allow state and regional data providers to validate the data at their convenience and frequency, without the burden of moving the data to an EPA platform. It will be designed to operate in environments where states have implemented SDWIS-STATE, or their own data management systems. It will take the EPA XML schema as input, and thus will take advantage of commercial off-the shelf (COTS) XML parser software for field and cross-field validations which precludes the need to develop custom software for that purpose.

Technology

SDWIS-FED
Data will flow from states to EPA in XML format, the current industry standard. The drinking water draft schema will be published in February 2003. The draft schema will be tested with several volunteer states' data. The staging tables on the NT-Server will accept data from any state ready to exchange XML formatted data through CDX as soon as the schema is judged ready.

ATTACHMENTS

Attachment 1: List of OGWDW Data Systems

#	Name	Acronym
List of current/planned information systems in OGWDW		
1	Safe Drinking Water Information System	SDWIS
2	National Contaminant Occurrence Database/Safe Drinking Water Accession and Review System	NCOD-SDWARS
3	Drinking Water Mapping Application	DWMA
4	Drinking Water National Information Management System	DWNIMS
5	Long-Term 2 Data Base	LT-2
6	National Environmental Method Index	NEMI
7	Drinking Water Research Database(?)	DRINK
8	Contaminant Information Tool	CIT
9	Contaminant Candidate List *(Planned)*	CCL